

TITLE
PLANT PROTEINASES

This application claims the benefit of U.S. Provisional Application No. 60/119,599, filed February 10, 1999.

FIELD OF THE INVENTION

This invention is in the field of plant molecular biology. More specifically, this invention pertains to nucleic acid fragments encoding proteinases in plants and seeds.

BACKGROUND OF THE INVENTION

The protein turnover in all organisms must be highly regulated since proteins within the same subcellular compartment have very different half-lives. The ubiquitin system is the major energy-dependent protease in the cytosol where ubiquitin binds to the protein to be degraded and then the multisubunit proteasome processes it. Another energy-dependent protease is the membrane-bound ATP-dependent proteinase or Clp. Other cellular proteinases are the calcium-dependent calpain, cysteine proteases, metallo-peptidases and other serine-type peptidases.

Calpain is an intracellular calcium-dependent protease activated at cell membranes which cleaves cytoskeletal and submembranous proteins and is probably involved in the calcium-dependent regulation of the cytoskeletal reorganization. The proteolytic activity of calpain and the binding of calpain to membranes is inhibited by the endogenous inhibitor calpastatin. Mammalian and insect tissues each has at least two different calpain isozymes. Each isozyme contains a different large subunit and an identical small subunit. A third calpain isozyme containing only a large subunit called p94 has been identified in mammalian tissues. All three large subunits contain four conserved domains including a cysteine protease domain and a calcium-binding domain. The p94 large subunit contains three additional domains probably involved in protease activity regulation and/or intracellular localization (Sorimachi et al. (1989) *J. Biol. Chem.* 264:20106-20111). While no calpain activity has been detected in chromatographic extracts from *Elodea densa* tissues (Wolfe et al (1989) *Life Sci.* 45:2093-2101), the p94 large subunit may be present at a lower concentration than the sensitivity of their assay (1 microgram per 0.25 g) and thus may be found in plant tissues.

Two different cysteine proteinases have been isolated from tissues of several species including arabidopsis, pea, rice, barley and corn. The barley cysteine proteinases EP-A and EP-B are induced by the presence of gibberellic acid and play a central role in the breakdown of endosperm storage proteins (hordeins) in the aleurone layers (Koehler and Ho (1988). *Plant Physiol.* 87:95-103). These cysteine proteinases are members of a small gene family composed of four to five different genes and are translated as prosequences which follow a post-translationally multistep processing to mature products. The barley cysteine proteinases are differentially hormonally induced and temporally regulated (Koehler and Ho

(1990) *Plant Cell* 2:769-783). Cysteine proteinases isolated from corn seeds are differentially expressed and have different subcellular localizations (Domoto et al. (1995) *Biochim. Biophys. Acta* 1263:241-244).

The CLP system was first identified in *E. coli* and later in plant chloroplasts. This proteolytic activity is induced by heat shock, by salt or oxidative stress or by glucose or oxygen limitation. Two different components are present in the CLP system, a protease and an ATP-binding factor. Two different types of ATP-binding factors, CLPA and CLPX, present specific substrates to the protease domain to facilitate their degradation. A six-membered ring formed by CLPX subunits binds to ATP and to two seven-member rings of CLPP to produce the active enzyme. This complex is structurally analogous to the one formed by CLPP and CLPA (Grimaud et al. (1998) *J. Biol. Chem.* 273:12476-12481).

SUMMARY OF THE INVENTION

The present invention relates to isolated polynucleotides comprising a nucleotide sequence encoding a polypeptide of at least 40 amino acids that has at least 85% identity based on the Clustal method of alignment when compared to a polypeptide selected from the group consisting of a calpain p94 polypeptide of SEQ ID NOs:2, 4, 6, 30, 32, and 34. The present invention also relates to an isolated polynucleotide comprising the complement of the nucleotide sequences described above.

The present invention relates to isolated polynucleotides comprising a nucleotide sequence encoding a polypeptide of at least 150 amino acids that has at least 95% identity based on the Clustal method of alignment when compared to a polypeptide selected from the group consisting of a cysteine protease 1 polypeptide of SEQ ID NOs:8, 10, 36, and 38. The present invention also relates to an isolated polynucleotide comprising the complement of the nucleotide sequences described above.

The present invention relates to isolated polynucleotides comprising a nucleotide sequence encoding a polypeptide of at least 200 amino acids that has at least 80% identity based on the Clustal method of alignment when compared to a polypeptide selected from the group consisting of a cysteine protease 2 polypeptide of SEQ ID NOs:12 and 40. The present invention also relates to an isolated polynucleotide comprising the complement of the nucleotide sequences described above.

The present invention relates to isolated polynucleotides comprising a nucleotide sequence encoding a polypeptide of at least 175 amino acids that has at least 95% identity based on the Clustal method of alignment when compared to a polypeptide selected from the group consisting of a CLPA polypeptide of SEQ ID NOs:14, 16, 18, 42, 44, and 46. The present invention also relates to an isolated polynucleotide comprising the complement of the nucleotide sequences described above.

The present invention relates to isolated polynucleotides comprising a nucleotide sequence encoding a CLPP polypeptide selected from the group consisting of SEQ ID

NOs:20, 22, 24, 26, 28, 48, 50, 52, 54, and 56. The present invention also relates to an isolated polynucleotide comprising the complement of the nucleotide sequences described above.

It is preferred that the isolated polynucleotides of the claimed invention consist of a nucleic acid sequence selected from the group consisting of SEQ ID NOs:1, 3, 5, 7, 9,11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, and 55 that codes for the polypeptide selected from the group consisting of SEQ ID NOs:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, and 56. The present invention also relates to an isolated polynucleotide comprising a nucleotide sequences of at least one of 60 (preferably at least one of 40, most preferably at least one of 30) contiguous nucleotides derived from a nucleotide sequence selected from the group consisting of SEQ ID NOs:1, 3, 5, 7, 9,11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, and 55 and the complement of such nucleotide sequences.

The present invention relates to a chimeric gene comprising an isolated polynucleotide of the present invention operably linked to suitable regulatory sequences.

The present invention relates to an isolated host cell comprising a chimeric gene of the present invention or an isolated polynucleotide of the present invention. The host cell may be eukaryotic, such as a yeast or a plant cell, or prokaryotic, such as a bacterial cell. The present invention also relates to a virus, preferably a baculovirus, comprising an isolated polynucleotide of the present invention or a chimeric gene of the present invention.

The present invention relates to a process for producing an isolated host cell comprising a chimeric gene of the present invention or an isolated polynucleotide of the present invention, the process comprising either transforming or transfecting an isolated compatible host cell with a chimeric gene or isolated polynucleotide of the present invention.

The present invention relates to a calpain p94 polypeptide of at least 40 amino acids comprising at least 85% homology based on the Clustal method of alignment compared to a polypeptide selected from the group consisting of SEQ ID NOs:2, 4, 6, 30, 32, and 34.

The present invention relates to a cysteine protease 1 polypeptide of at least 150 amino acids comprising at least 95% homology based on the Clustal method of alignment compared to a polypeptide selected from the group consisting of SEQ ID NOs:8, 10, 36, and 38.

The present invention relates to a cysteine protease 2 polypeptide of at least 200 amino acids comprising at least 80% homology based on the Clustal method of alignment compared to a polypeptide selected from the group consisting of SEQ ID NOs:12 and 40.

The present invention relates to a CLPA polypeptide of at least 175 amino acids comprising at least 95% homology based on the Clustal method of alignment compared to a polypeptide selected from the group consisting of SEQ ID NOs:14, 16, 18, 42, 44, and 46.

The present invention relates to a CLPP polypeptide selected from the group consisting of SEQ ID NOs:20, 22, 24, 26, 28, 48, 50, 52, 54, and 56.

The present invention relates to a method of selecting an isolated polynucleotide that affects the level of expression of a proteinase polypeptide in a host cell, preferably a plant cell, the method comprising the steps of: (a) constructing an isolated polynucleotide of the present invention or an isolated chimeric gene of the present invention; (b) introducing the isolated polynucleotide or the isolated chimeric gene into a host cell; (c) measuring the level the cysteine protease I, the cysteine protase 2, the calpain large subunit, the CLP protease proteolytic subunit or the CLP protease ATP binding subunit polypeptide in the host cell containing the isolated polynucleotide; and (d) comparing the level of the cysteine protease I, the cysteine protase 2, the calpain large subunit, the CLP protease proteolytic subunit or the CLP protease ATP binding subunit polypeptide in the host cell containing the isolated polynucleotide with the level of the cysteine protease I, the cysteine protase 2, the calpain large subunit, the CLP protease proteolytic subunit or the CLP protease ATP binding subunit polypeptide in the host cell that does not contain the isolated polynucleotide.

The present invention relates to a method of obtaining a nucleic acid fragment encoding a substantial portion of a cysteine protease I, a cysteine protase 2, a calpain large subunit, a CLP protease proteolytic subunit or a CLP protease ATP binding subunit polypeptide, preferably a plant cysteine protease I, a cysteine protase 2, a calpain large subunit, a CLP protease proteolytic subunit or a CLP protease ATP binding subunit polypeptide, comprising the steps of : synthesizing an oligonucleotide primer comprising a nucleotide sequence of at least one of 60 (preferably at least one of 40, most preferably at least one of 30) contiguous nucleotides derived from a nucleotide sequence selected from the group consisting of SEQ ID NOs:1, 3, 5, 7, 9,11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, and 55 and the complement of such nucleotide sequences; and amplifying a nucleic acid fragment (preferably a cDNA inserted in a cloning vector) using the oligonucleotide primer. The amplified nucleic acid fragment preferably will encode a portion of a cysteine protease I, a cysteine protase 2, a calpain large subunit, a CLP protease proteolytic subunit or a CLP protease ATP binding subunit amino acid sequence.

The present invention also relates to a method of obtaining a nucleic acid fragment encoding all or a substantial portion of the amino acid sequence encoding a cysteine protease I, a cysteine protase 2, a calpain large subunit, a CLP protease proteolytic subunit or a CLP protease ATP binding subunit polypeptide comprising the steps of: probing a cDNA or genomic library with an isolated polynucleotide of the present invention; identifying a DNA clone that hybridizes with an isolated polynucleotide of the present invention; isolating the identified DNA clone; and sequencing the cDNA or genomic fragment that comprises the isolated DNA clone.

The present invention relates to a composition, such as a hybridization mixture, comprising an isolated polynucleotide of the present invention.

The present invention relates to an isolated polynucleotide of the present invention comprising at least one of 30 contiguous nucleotides derived from a nucleic acid sequence selected from the group consisting of SEQ ID NOs:1, 3, 5, 7, 9,11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, and 55.

The present invention relates to an expression cassette comprising an isolated polynucleotide of the present invention operably linked to a promoter.

The present invention relates to a method for positive selection of a transformed cell comprising: (a) transforming a host cell with the chimeric gene of the present invention or an expression cassette of the present invention; and (b) growing the transformed host cell, preferably plant cell, such as a monocot or a dicot, under conditions which allow expression of the cysteine protease I, the cysteine protease 2, the calpain large subunit, the CLP protease proteolytic subunit or the CLP protease ATP binding subunit polynucleotide in an amount sufficient to complement a null mutant to provide a positive selection means.

BRIEF DESCRIPTION OF THE SEQUENCE LISTINGS

The invention can be more fully understood from the following detailed description and the accompanying Sequence Listing which form a part of this application.

Table 1 lists the polypeptides that are described herein, the designation of the cDNA clones that comprise the nucleic acid fragments encoding polypeptides representing all or a substantial portion of these polypeptides, and the corresponding identifier (SEQ ID NO:) as used in the attached Sequence Listing. The sequence descriptions and Sequence Listing attached hereto comply with the rules governing nucleotide and/or amino acid sequence disclosures in patent applications as set forth in 37 C.F.R. §1.821-1.825.

TABLE 1
PLANT PROTEINASES

Protein	Clone Designation	SEQ ID NO:	
		(Nucleotide)	(Amino Acid)
Corn Calpain p94 Subunit	cbn2.pk0039.c2	1	2
Rice Calpain p94 Subunit	rs11n.pk013.h14	3	4
Soybean Calpain p94 Subunit	ses9c.pk001.j23	5	6
Rice Cysteine Protease 1	rr1.pk084.j16	7	8
Wheat Cysteine Protease 1	Contig of: wdk1c.pk009.j19 wre1n.pk164.b11	9	10
Soybean Cysteine Protease 2	Contig of: sgs2c.pk002.p14 sr3c.pk003.d10 scb1c.pk003.d8	11	12

Protein	Clone Designation	SEQ ID NO:	
		(Nucleotide)	(Amino Acid)
Corn CLP ATP Binding Subunit	p0110.cgsmk69r	13	14
Rice CLP ATP Binding Subunit	Contig of: rlr6.pk0083.f9 rlr24.pk0088.f7 rlr6.pk0029.d7	15	16
Wheat CLP ATP Binding Subunit	wlm96.pk032.n8	17	18
Corn CLP Proteolytic Subunit	p0060.coran66r	19	20
Rice CLP Proteolytic Subunit	rsr9n.pk004.p5	21	22
Soybean CLP Proteolytic Subunit	scb1c.pk004.k24	23	24
Wheat CLP Proteolytic Subunit	wle1n.pk0042.f7	25	26
Wheat CLP Proteolytic Subunit	wlk8.pk0006.a4	27	28
Corn Calpain p94 Subunit	cbn2.pk0039.c2:fis	29	30
Rice Calpain p94 Subunit	rs11n.pk013.h14:fis	31	32
Soybean Calpain p94 Subunit	ses9c.pk001.j23:fis	33	34
Rice Cysteine Protease 1	rr1.pk084.j16:fis	35	36
Wheat Cysteine Protease 1	wdk1c.pk009.j19:fis	37	38
Soybean Cysteine Protease 2	srr3c.pk003.d10:fis	39	40
Corn CLP ATP Binding Subunit	p0110.cgsmk69r:fis	41	42
Rice CLP ATP Binding Subunit	rlr24.pk0088.f7:fis	43	44
Wheat CLP ATP Binding Subunit	wlm96.pk032.n8:fis	45	46
Corn CLP Proteolytic Subunit	p0060.coran66r:fis	47	48
Rice CLP Proteolytic Subunit	rsr9n.pk004.p5:fis	49	50
Soybean CLP Proteolytic Subunit	scb1c.pk004.k24:fis	51	52
Wheat CLP Proteolytic Subunit	wle1n.pk0042.f7:fis	53	54
Wheat CLP Proteolytic Subunit	wlk8.pk0006.a4:fis	55	56

The Sequence Listing contains the one letter code for nucleotide sequence characters and the three letter codes for amino acids as defined in conformity with the IUPAC-IUBMB standards described in *Nucleic Acids Res.* 13:3021-3030 (1985) and in the *Biochemical J.* 219 (No. 2):345-373 (1984) which are herein incorporated by reference. The symbols and format used for nucleotide and amino acid sequence data comply with the rules set forth in 37 C.F.R. §1.822.

DETAILED DESCRIPTION OF THE INVENTION

In the context of this disclosure, a number of terms shall be utilized. As used herein, a "polynucleotide" is a nucleotide sequence such as a nucleic acid fragment. A polynucleotide may be a polymer of RNA or DNA that is single- or double-stranded, that optionally

contains synthetic, non-natural or altered nucleotide bases. A polynucleotide in the form of a polymer of DNA may be comprised of one or more segments of cDNA, genomic DNA, synthetic DNA, or mixtures thereof. An isolated polynucleotide of the present invention may include at least one of 60 contiguous nucleotides, preferably at least one of 40

5 contiguous nucleotides, most preferably one of at least 30 contiguous nucleotides derived from SEQ ID NOs:1, 3, 5, 7, 9,11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, and 55, or the complement of such sequences.

As used herein, "contig" refers to a nucleotide sequence that is assembled from two or more constituent nucleotide sequences that share common or overlapping regions of
10 sequence homology. For example, the nucleotide sequences of two or more nucleic acid fragments can be compared and aligned in order to identify common or overlapping sequences. Where common or overlapping sequences exist between two or more nucleic acid fragments, the sequences (and thus their corresponding nucleic acid fragments) can be assembled into a single contiguous nucleotide sequence.

15 As used herein, "substantially similar" refers to nucleic acid fragments wherein changes in one or more nucleotide bases results in substitution of one or more amino acids, but do not affect the functional properties of the polypeptide encoded by the nucleotide sequence. "Substantially similar" also refers to nucleic acid fragments wherein changes in one or more nucleotide bases does not affect the ability of the nucleic acid fragment to
20 mediate alteration of gene expression by gene silencing through for example antisense or co-suppression technology. "Substantially similar" also refers to modifications of the nucleic acid fragments of the instant invention such as deletion or insertion of one or more nucleotides that do not substantially affect the functional properties of the resulting transcript vis-à-vis the ability to mediate gene silencing or alteration of the functional
25 properties of the resulting protein molecule. It is therefore understood that the invention encompasses more than the specific exemplary nucleotide or amino acid sequences and includes functional equivalents thereof.

Substantially similar nucleic acid fragments may be selected by screening nucleic acid fragments representing subfragments or modifications of the nucleic acid fragments of the
30 instant invention, wherein one or more nucleotides are substituted, deleted and/or inserted, for their ability to affect the level of the polypeptide encoded by the unmodified nucleic acid fragment in a plant or plant cell. For example, a substantially similar nucleic acid fragment representing at least one of 30 contiguous nucleotides derived from the instant nucleic acid fragment can be constructed and introduced into a plant or plant cell. The level of the
35 polypeptide encoded by the unmodified nucleic acid fragment present in a plant or plant cell exposed to the substantially similar nucleic fragment can then be compared to the level of the polypeptide in a plant or plant cell that is not exposed to the substantially similar nucleic acid fragment.

For example, it is well known in the art that antisense suppression and co-suppression of gene expression may be accomplished using nucleic acid fragments representing less than the entire coding region of a gene, and by nucleic acid fragments that do not share 100% sequence identity with the gene to be suppressed. Moreover, alterations in a nucleic acid fragment which result in the production of a chemically equivalent amino acid at a given site, but do not effect the functional properties of the encoded polypeptide, are well known in the art. Thus, a codon for the amino acid alanine, a hydrophobic amino acid, may be substituted by a codon encoding another less hydrophobic residue, such as glycine, or a more hydrophobic residue, such as valine, leucine, or isoleucine. Similarly, changes which result in substitution of one negatively charged residue for another, such as aspartic acid for glutamic acid, or one positively charged residue for another, such as lysine for arginine, can also be expected to produce a functionally equivalent product. Nucleotide changes which result in alteration of the N-terminal and C-terminal portions of the polypeptide molecule would also not be expected to alter the activity of the polypeptide. Each of the proposed modifications is well within the routine skill in the art, as is determination of retention of biological activity of the encoded products. Consequently, an isolated polynucleotide comprising a nucleotide sequence of at least one of 60 (preferably at least one of 40, most preferably at least one of 30) contiguous nucleotides derived from a nucleotide sequence selected from the group consisting of SEQ ID NOs:1, 3, 5, 7, 9,11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, and 55 and the complement of such nucleotide sequences may be used in methods of selecting an isolated polynucleotide that affects the expression of a proteinase polypeptide in a host cell. A method of selecting an isolated polynucleotide that affects the level of expression of a polypeptide in a host cell (eukaryotic, such as plant or yeast, prokaryotic such as bacterial, or viral) may comprise the steps of: constructing an isolated polynucleotide of the present invention or an isolated chimeric gene of the present invention; introducing the isolated polynucleotide or the isolated chimeric gene into a host cell; measuring the level a polypeptide in the host cell containing the isolated polynucleotide; and comparing the level of a polypeptide in the host cell containing the isolated polynucleotide with the level of a polypeptide in a host cell that does not contain the isolated polynucleotide.

Moreover, substantially similar nucleic acid fragments may also be characterized by their ability to hybridize. Estimates of such homology are provided by either DNA-DNA or DNA-RNA hybridization under conditions of stringency as is well understood by those skilled in the art (Hames and Higgins, Eds. (1985) *Nucleic Acid Hybridisation*, IRL Press, Oxford, U.K.). Stringency conditions can be adjusted to screen for moderately similar fragments, such as homologous sequences from distantly related organisms, to highly similar fragments, such as genes that duplicate functional enzymes from closely related organisms. Post-hybridization washes determine stringency conditions. One set of preferred conditions

uses a series of washes starting with 6X SSC, 0.5% SDS at room temperature for 15 min, then repeated with 2X SSC, 0.5% SDS at 45°C for 30 min, and then repeated twice with 0.2X SSC, 0.5% SDS at 50°C for 30 min. A more preferred set of stringent conditions uses higher temperatures in which the washes are identical to those above except for the temperature of the final two 30 min washes in 0.2X SSC, 0.5% SDS was increased to 60°C. Another preferred set of highly stringent conditions uses two final washes in 0.1X SSC, 0.1% SDS at 65°C.

Substantially similar nucleic acid fragments of the instant invention may also be characterized by the percent identity of the amino acid sequences that they encode to the amino acid sequences disclosed herein, as determined by algorithms commonly employed by those skilled in this art. Suitable nucleic acid fragments (isolated polynucleotides of the present invention) encode polypeptides that are at least about 70% identical, preferably at least about 80% identical to the amino acid sequences reported herein. Preferred nucleic acid fragments encode amino acid sequences that are about 85% identical to the amino acid sequences reported herein. More preferred nucleic acid fragments encode amino acid sequences that are at least about 90% identical to the amino acid sequences reported herein. Most preferred are nucleic acid fragments that encode amino acid sequences that are at least about 95% identical to the amino acid sequences reported herein. Suitable nucleic acid fragments not only have the above homologies but typically encode a polypeptide having at least 50 amino acids, preferably at least 100 amino acids, more preferably at least 150 amino acids, still more preferably at least 200 amino acids, and most preferably at least 250 amino acids. Sequence alignments and percent identity calculations were performed using the Megalign program of the LASERGENE bioinformatics computing suite (DNASTAR Inc., Madison, WI). Multiple alignment of the sequences was performed using the Clustal method of alignment (Higgins and Sharp (1989) *CABIOS*. 5:151-153) with the default parameters (GAP PENALTY=10, GAP LENGTH PENALTY=10). Default parameters for pairwise alignments using the Clustal method were KTUPLE 1, GAP PENALTY=3, WINDOW=5 and DIAGONALS SAVED=5.

A "substantial portion" of an amino acid or nucleotide sequence comprises an amino acid or a nucleotide sequence that is sufficient to afford putative identification of the protein or gene that the amino acid or nucleotide sequence comprises. Amino acid and nucleotide sequences can be evaluated either manually by one skilled in the art, or by using computer-based sequence comparison and identification tools that employ algorithms such as BLAST (Basic Local Alignment Search Tool; Altschul et al. (1993) *J. Mol. Biol.* 215:403-410; see also www.ncbi.nlm.nih.gov/BLAST/). In general, a sequence of ten or more contiguous amino acids or thirty or more contiguous nucleotides is necessary in order to putatively identify a polypeptide or nucleic acid sequence as homologous to a known protein or gene. Moreover, with respect to nucleotide sequences, gene-specific oligonucleotide probes

comprising 30 or more contiguous nucleotides may be used in sequence-dependent methods of gene identification (e.g., Southern hybridization) and isolation (e.g., *in situ* hybridization of bacterial colonies or bacteriophage plaques). In addition, short oligonucleotides of 12 or more nucleotides may be used as amplification primers in PCR in order to obtain a particular nucleic acid fragment comprising the primers. Accordingly, a “substantial portion” of a nucleotide sequence comprises a nucleotide sequence that will afford specific identification and/or isolation of a nucleic acid fragment comprising the sequence. The instant specification teaches amino acid and nucleotide sequences encoding polypeptides that comprise one or more particular plant proteins. The skilled artisan, having the benefit of the sequences as reported herein, may now use all or a substantial portion of the disclosed sequences for purposes known to those skilled in this art. Accordingly, the instant invention comprises the complete sequences as reported in the accompanying Sequence Listing, as well as substantial portions of those sequences as defined above.

“Codon degeneracy” refers to divergence in the genetic code permitting variation of the nucleotide sequence without effecting the amino acid sequence of an encoded polypeptide. Accordingly, the instant invention relates to any nucleic acid fragment comprising a nucleotide sequence that encodes all or a substantial portion of the amino acid sequences set forth herein. The skilled artisan is well aware of the “codon-bias” exhibited by a specific host cell in usage of nucleotide codons to specify a given amino acid. Therefore, when synthesizing a nucleic acid fragment for improved expression in a host cell, it is desirable to design the nucleic acid fragment such that its frequency of codon usage approaches the frequency of preferred codon usage of the host cell.

“Synthetic nucleic acid fragments” can be assembled from oligonucleotide building blocks that are chemically synthesized using procedures known to those skilled in the art. These building blocks are ligated and annealed to form larger nucleic acid fragments which may then be enzymatically assembled to construct the entire desired nucleic acid fragment. “Chemically synthesized”, as related to nucleic acid fragment, means that the component nucleotides were assembled *in vitro*. Manual chemical synthesis of nucleic acid fragments may be accomplished using well established procedures, or automated chemical synthesis can be performed using one of a number of commercially available machines. Accordingly, the nucleic acid fragments can be tailored for optimal gene expression based on optimization of nucleotide sequence to reflect the codon bias of the host cell. The skilled artisan appreciates the likelihood of successful gene expression if codon usage is biased towards those codons favored by the host. Determination of preferred codons can be based on a survey of genes derived from the host cell where sequence information is available.

“Gene” refers to a nucleic acid fragment that expresses a specific protein, including regulatory sequences preceding (5' non-coding sequences) and following (3' non-coding sequences) the coding sequence. “Native gene” refers to a gene as found in nature with its

own regulatory sequences. “Chimeric gene” refers any gene that is not a native gene, comprising regulatory and coding sequences that are not found together in nature.

Accordingly, a chimeric gene may comprise regulatory sequences and coding sequences that are derived from different sources, or regulatory sequences and coding sequences derived

5 from the same source, but arranged in a manner different than that found in nature.

“Endogenous gene” refers to a native gene in its natural location in the genome of an organism. A “foreign” gene refers to a gene not normally found in the host organism, but that is introduced into the host organism by gene transfer. Foreign genes can comprise native genes inserted into a non-native organism, or chimeric genes. A “transgene” is a gene

10 that has been introduced into the genome by a transformation procedure.

“Coding sequence” refers to a nucleotide sequence that codes for a specific amino acid sequence. “Regulatory sequences” refer to nucleotide sequences located upstream (5' non-coding sequences), within, or downstream (3' non-coding sequences) of a coding sequence, and which influence the transcription, RNA processing or stability, or translation of the

15 associated coding sequence. Regulatory sequences may include promoters, translation leader sequences, introns, and polyadenylation recognition sequences.

“Promoter” refers to a nucleotide sequence capable of controlling the expression of a coding sequence or functional RNA. In general, a coding sequence is located 3' to a promoter sequence. The promoter sequence consists of proximal and more distal upstream elements, the latter elements often referred to as enhancers. Accordingly, an “enhancer” is a nucleotide sequence which can stimulate promoter activity and may be an innate element of the promoter or a heterologous element inserted to enhance the level or tissue-specificity of a promoter. Promoters may be derived in their entirety from a native gene, or be composed

25 of synthetic nucleotide segments. It is understood by those skilled in the art that different promoters may direct the expression of a gene in different tissues or cell types, or at different stages of development, or in response to different environmental conditions. Promoters which cause a nucleic acid fragment to be expressed in most cell types at most times are commonly referred to as “constitutive promoters”. New promoters of various

30 types useful in plant cells are constantly being discovered; numerous examples may be found in the compilation by Okamuro and Goldberg (1989) *Biochemistry of Plants* 15:1-82. It is further recognized that since in most cases the exact boundaries of regulatory sequences have not been completely defined, nucleic acid fragments of different lengths may have identical promoter activity.

35 The “translation leader sequence” refers to a nucleotide sequence located between the promoter sequence of a gene and the coding sequence. The translation leader sequence is present in the fully processed mRNA upstream of the translation start sequence. The translation leader sequence may affect processing of the primary transcript to mRNA,

mRNA stability or translation efficiency. Examples of translation leader sequences have been described (Turner and Foster (1995) *Mol. Biotechnol.* 3:225-236).

The “3' non-coding sequences” refer to nucleotide sequences located downstream of a coding sequence and include polyadenylation recognition sequences and other sequences encoding regulatory signals capable of affecting mRNA processing or gene expression. The polyadenylation signal is usually characterized by affecting the addition of polyadenylic acid tracts to the 3' end of the mRNA precursor. The use of different 3' non-coding sequences is exemplified by Ingelbrecht et al. (1989) *Plant Cell* 1:671-680.

“RNA transcript” refers to the product resulting from RNA polymerase-catalyzed transcription of a DNA sequence. When the RNA transcript is a perfect complementary copy of the DNA sequence, it is referred to as the primary transcript or it may be a RNA sequence derived from posttranscriptional processing of the primary transcript and is referred to as the mature RNA. “Messenger RNA (mRNA)” refers to the RNA that is without introns and that can be translated into polypeptide by the cell. “cDNA” refers to a double-stranded DNA that is complementary to and derived from mRNA. “Sense” RNA refers to an RNA transcript that includes the mRNA and so can be translated into a polypeptide by the cell. “Antisense RNA” refers to an RNA transcript that is complementary to all or part of a target primary transcript or mRNA and that blocks the expression of a target gene (see U.S. Patent No. 5,107,065, incorporated herein by reference). The complementarity of an antisense RNA may be with any part of the specific nucleotide sequence, i.e., at the 5' non-coding sequence, 3' non-coding sequence, introns, or the coding sequence. “Functional RNA” refers to sense RNA, antisense RNA, ribozyme RNA, or other RNA that may not be translated but yet has an effect on cellular processes.

The term “operably linked” refers to the association of two or more nucleic acid fragments on a single nucleic acid fragment so that the function of one is affected by the other. For example, a promoter is operably linked with a coding sequence when it is capable of affecting the expression of that coding sequence (i.e., that the coding sequence is under the transcriptional control of the promoter). Coding sequences can be operably linked to regulatory sequences in sense or antisense orientation.

The term “expression”, as used herein, refers to the transcription and stable accumulation of sense (mRNA) or antisense RNA derived from the nucleic acid fragment of the invention. Expression may also refer to translation of mRNA into a polypeptide. “Antisense inhibition” refers to the production of antisense RNA transcripts capable of suppressing the expression of the target protein. “Overexpression” refers to the production of a gene product in transgenic organisms that exceeds levels of production in normal or non-transformed organisms. “Co-suppression” refers to the production of sense RNA transcripts capable of suppressing the expression of identical or substantially similar foreign or endogenous genes (U.S. Patent No. 5,231,020, incorporated herein by reference).

“Altered levels” refers to the production of gene product(s) in transgenic organisms in amounts or proportions that differ from that of normal or non-transformed organisms.

“Mature” protein refers to a post-translationally processed polypeptide; i.e., one from which any pre- or propeptides present in the primary translation product have been removed.

5 “Precursor” protein refers to the primary product of translation of mRNA; i.e., with pre- and propeptides still present. Pre- and propeptides may be but are not limited to intracellular localization signals.

10 A “chloroplast transit peptide” is an amino acid sequence which is translated in conjunction with a protein and directs the protein to the chloroplast or other plastid types present in the cell in which the protein is made. “Chloroplast transit sequence” refers to a nucleotide sequence that encodes a chloroplast transit peptide. A “signal peptide” is an amino acid sequence which is translated in conjunction with a protein and directs the protein to the secretory system (Chrispeels (1991) *Ann. Rev. Plant Phys. Plant Mol. Biol.* 42:21-53). If the protein is to be directed to a vacuole, a vacuolar targeting signal (*supra*) can further be
15 added, or if to the endoplasmic reticulum, an endoplasmic reticulum retention signal (*supra*) may be added. If the protein is to be directed to the nucleus, any signal peptide present should be removed and instead a nuclear localization signal included (Raikhel (1992) *Plant Phys.* 100:1627-1632).

20 “Transformation” refers to the transfer of a nucleic acid fragment into the genome of a host organism, resulting in genetically stable inheritance. Host organisms containing the transformed nucleic acid fragments are referred to as “transgenic” organisms. Examples of methods of plant transformation include *Agrobacterium*-mediated transformation (De Blaere et al. (1987) *Meth. Enzymol.* 143:277) and particle-accelerated or “gene gun” transformation technology (Klein et al. (1987) *Nature (London)* 327:70-73; U.S. Patent No. 4,945,050,
25 incorporated herein by reference).

Standard recombinant DNA and molecular cloning techniques used herein are well known in the art and are described more fully in Sambrook et al. *Molecular Cloning: A Laboratory Manual*; Cold Spring Harbor Laboratory Press: Cold Spring Harbor, 1989 (hereinafter “Maniatis”).

30 Nucleic acid fragments encoding at least a portion of several proteinases have been isolated and identified by comparison of random plant cDNA sequences to public databases containing nucleotide and protein sequences using the BLAST algorithms well known to those skilled in the art. The nucleic acid fragments of the instant invention may be used to isolate cDNAs and genes encoding homologous proteins from the same or other plant
35 species. Isolation of homologous genes using sequence-dependent protocols is well known in the art. Examples of sequence-dependent protocols include, but are not limited to, methods of nucleic acid hybridization, and methods of DNA and RNA amplification as

exemplified by various uses of nucleic acid amplification technologies (e.g., polymerase chain reaction, ligase chain reaction).

For example, genes encoding other cysteine protease 1s, cysteine protease 2s, calpain p94s, CLPAs, or CLPPs, either as cDNAs or genomic DNAs, could be isolated directly by using all or a portion of the instant nucleic acid fragments as DNA hybridization probes to screen libraries from any desired plant employing methodology well known to those skilled in the art. Specific oligonucleotide probes based upon the instant nucleic acid sequences can be designed and synthesized by methods known in the art (Maniatis). Moreover, the entire sequences can be used directly to synthesize DNA probes by methods known to the skilled artisan such as random primer DNA labeling, nick translation, or end-labeling techniques, or RNA probes using available *in vitro* transcription systems. In addition, specific primers can be designed and used to amplify a part or all of the instant sequences. The resulting amplification products can be labeled directly during amplification reactions or labeled after amplification reactions, and used as probes to isolate full length cDNA or genomic fragments under conditions of appropriate stringency.

In addition, two short segments of the instant nucleic acid fragments may be used in polymerase chain reaction protocols to amplify longer nucleic acid fragments encoding homologous genes from DNA or RNA. The polymerase chain reaction may also be performed on a library of cloned nucleic acid fragments wherein the sequence of one primer is derived from the instant nucleic acid fragments, and the sequence of the other primer takes advantage of the presence of the polyadenylic acid tracts to the 3' end of the mRNA precursor encoding plant genes. Alternatively, the second primer sequence may be based upon sequences derived from the cloning vector. For example, the skilled artisan can follow the RACE protocol (Frohman et al. (1988) *Proc. Natl. Acad. Sci. USA* 85:8998-9002) to generate cDNAs by using PCR to amplify copies of the region between a single point in the transcript and the 3' or 5' end. Primers oriented in the 3' and 5' directions can be designed from the instant sequences. Using commercially available 3' RACE or 5' RACE systems (BRL), specific 3' or 5' cDNA fragments can be isolated (Ohara et al. (1989) *Proc. Natl. Acad. Sci. USA* 86:5673-5677; Loh et al. (1989) *Science* 243:217-220). Products generated by the 3' and 5' RACE procedures can be combined to generate full-length cDNAs (Frohman and Martin (1989) *Techniques* 1:165). Consequently, a polynucleotide comprising a nucleotide sequence of at least one of 60 (preferably one of at least 40, most preferably one of at least 30) contiguous nucleotides derived from a nucleotide sequence selected from the group consisting of SEQ ID NOs:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, and 55 and the complement of such nucleotide sequences may be used in such methods to obtain a nucleic acid fragment encoding a substantial portion of an amino acid sequence of a polypeptide. The present invention relates to a method of obtaining a nucleic acid fragment encoding a substantial portion of a

proteinase polypeptide (such as a cysteine protease 1, a cysteine protease 2, a calpain large subunit, a CLP protease proteolytic subunit or a CLP protease ATP binding subunit) preferably a substantial portion of a plant polypeptide of a gene, comprising the steps of : synthesizing an oligonucleotide primer comprising a nucleotide sequence of at least one of 5 60 (preferably at least one of 40, most preferably at least one of 30) contiguous nucleotides derived from a nucleotide sequence selected from the group consisting of SEQ ID NOs:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, and 55, and the complement of such nucleotide sequences; and amplifying a nucleic acid fragment (preferably a cDNA inserted in a cloning vector) using the oligonucleotide primer. 10 The amplified nucleic acid fragment preferably will encode a portion of a cysteine protease I, a cysteine protease 2, a calpain large subunit, a CLP protease proteolytic subunit or a CLP protease ATP binding subunit polypeptide.

Availability of the instant nucleotide and deduced amino acid sequences facilitates immunological screening of cDNA expression libraries. Synthetic peptides representing 15 portions of the instant amino acid sequences may be synthesized. These peptides can be used to immunize animals to produce polyclonal or monoclonal antibodies with specificity for peptides or proteins comprising the amino acid sequences. These antibodies can be then be used to screen cDNA expression libraries to isolate full-length cDNA clones of interest (Lerner (1984) *Adv. Immunol.* 36:1-34; Maniatis).

20 The nucleic acid fragments of the instant invention may be used to create transgenic plants in which the disclosed polypeptides are present at higher or lower levels than normal or in cell types or developmental stages in which they are not normally found. This would have the effect of altering the level of protein stability in those cells. The cysteine proteinases or the calpain p94 subunit are useful for the study and control of apoptosis in 25 plants since they are induced by stress. Manipulating the expression of proteinases in plant cells will be useful for controlling cell death (apoptosis) caused by disease. Manipulation of the expression of proteinases in tissue culture will improve the survival rate during the harsh transformation treatments and during maintenance in tissue culture.

Overexpression of the proteins of the instant invention may be accomplished by first 30 constructing a chimeric gene in which the coding region is operably linked to a promoter capable of directing expression of a gene in the desired tissues at the desired stage of development. The chimeric gene may comprise promoter sequences and translation leader sequences derived from the same genes. 3' Non-coding sequences encoding transcription termination signals may also be provided. The instant chimeric gene may also comprise one 35 or more introns in order to facilitate gene expression.

Plasmid vectors comprising the isolated polynucleotide (or chimeric gene) may be constructed. The choice of plasmid vector is dependent upon the method that will be used to transform host plants. The skilled artisan is well aware of the genetic elements that must be

present on the plasmid vector in order to successfully transform, select and propagate host cells containing the chimeric gene. The skilled artisan will also recognize that different independent transformation events will result in different levels and patterns of expression (Jones et al. (1985) *EMBO J.* 4:2411-2418; De Almeida et al. (1989) *Mol. Gen. Genetics* 218:78-86), and thus that multiple events must be screened in order to obtain lines displaying the desired expression level and pattern. Such screening may be accomplished by Southern analysis of DNA, Northern analysis of mRNA expression, Western analysis of protein expression, or phenotypic analysis.

For some applications it may be useful to direct the instant polypeptides to different cellular compartments, or to facilitate its secretion from the cell. It is thus envisioned that the chimeric gene described above may be further supplemented by directing the coding sequence to encode the instant polypeptides with appropriate intracellular targeting sequences such as transit sequences (Keegstra (1989) *Cell* 56:247-253), signal sequences or sequences encoding endoplasmic reticulum localization (Chrispeels (1991) *Ann. Rev. Plant Phys. Plant Mol. Biol.* 42:21-53), or nuclear localization signals (Raikhel (1992) *Plant Phys.* 100:1627-1632) with or without removing targeting sequences that are already present. While the references cited give examples of each of these, the list is not exhaustive and more targeting signals of use may be discovered in the future.

It may also be desirable to reduce or eliminate expression of genes encoding the instant polypeptides in plants for some applications. In order to accomplish this, a chimeric gene designed for co-suppression of the instant polypeptide can be constructed by linking a gene or gene fragment encoding that polypeptide to plant promoter sequences. Alternatively, a chimeric gene designed to express antisense RNA for all or part of the instant nucleic acid fragment can be constructed by linking the gene or gene fragment in reverse orientation to plant promoter sequences. Either the co-suppression or antisense chimeric genes could be introduced into plants via transformation wherein expression of the corresponding endogenous genes are reduced or eliminated.

Molecular genetic solutions to the generation of plants with altered gene expression have a decided advantage over more traditional plant breeding approaches. Changes in plant phenotypes can be produced by specifically inhibiting expression of one or more genes by antisense inhibition or cosuppression (U.S. Patent Nos. 5,190,931, 5,107,065 and 5,283,323). An antisense or cosuppression construct would act as a dominant negative regulator of gene activity. While conventional mutations can yield negative regulation of gene activity these effects are most likely recessive. The dominant negative regulation available with a transgenic approach may be advantageous from a breeding perspective. In addition, the ability to restrict the expression of specific phenotype to the reproductive tissues of the plant by the use of tissue specific promoters may confer agronomic

advantages relative to conventional mutations which may have an effect in all tissues in which a mutant gene is ordinarily expressed.

The person skilled in the art will know that special considerations are associated with the use of antisense or cosuppression technologies in order to reduce expression of particular genes. For example, the proper level of expression of sense or antisense genes may require the use of different chimeric genes utilizing different regulatory elements known to the skilled artisan. Once transgenic plants are obtained by one of the methods described above, it will be necessary to screen individual transgenics for those that most effectively display the desired phenotype. Accordingly, the skilled artisan will develop methods for screening large numbers of transformants. The nature of these screens will generally be chosen on practical grounds. For example, one can screen by looking for changes in gene expression by using antibodies specific for the protein encoded by the gene being suppressed, or one could establish assays that specifically measure enzyme activity. A preferred method will be one which allows large numbers of samples to be processed rapidly, since it will be expected that a large number of transformants will be negative for the desired phenotype.

The instant polypeptides (or portions thereof) may be produced in heterologous host cells, particularly in the cells of microbial hosts, and can be used to prepare antibodies to the these proteins by methods well known to those skilled in the art. The antibodies are useful for detecting the polypeptides of the instant invention *in situ* in cells or *in vitro* in cell extracts. Preferred heterologous host cells for production of the instant polypeptides are microbial hosts. Microbial expression systems and expression vectors containing regulatory sequences that direct high level expression of foreign proteins are well known to those skilled in the art. Any of these could be used to construct a chimeric gene for production of the instant polypeptides. This chimeric gene could then be introduced into appropriate microorganisms via transformation to provide high level expression of the encoded proteinases. An example of a vector for high level expression of the instant polypeptides in a bacterial host is provided (Example 10).

All or a substantial portion of the nucleic acid fragments of the instant invention may also be used as probes for genetically and physically mapping the genes that they are a part of, and as markers for traits linked to those genes. Such information may be useful in plant breeding in order to develop lines with desired phenotypes. For example, the instant nucleic acid fragments may be used as restriction fragment length polymorphism (RFLP) markers. Southern blots (Maniatis) of restriction-digested plant genomic DNA may be probed with the nucleic acid fragments of the instant invention. The resulting banding patterns may then be subjected to genetic analyses using computer programs such as MapMaker (Lander et al. (1987) *Genomics* 1:174-181) in order to construct a genetic map. In addition, the nucleic acid fragments of the instant invention may be used to probe Southern blots containing restriction endonuclease-treated genomic DNAs of a set of individuals representing parent

and progeny of a defined genetic cross. Segregation of the DNA polymorphisms is noted and used to calculate the position of the instant nucleic acid sequence in the genetic map previously obtained using this population (Botstein et al. (1980) *Am. J. Hum. Genet.* 32:314-331).

5 The production and use of plant gene-derived probes for use in genetic mapping is described in Bernatzky and Tanksley (1986) *Plant Mol. Biol. Reporter* 4:37-41. Numerous publications describe genetic mapping of specific cDNA clones using the methodology outlined above or variations thereof. For example, F2 intercross populations, backcross populations, randomly mated populations, near isogenic lines, and other sets of individuals
10 may be used for mapping. Such methodologies are well known to those skilled in the art.

Nucleic acid probes derived from the instant nucleic acid sequences may also be used for physical mapping (i.e., placement of sequences on physical maps; *see* Hoheisel et al. In: *Nonmammalian Genomic Analysis: A Practical Guide*, Academic press 1996, pp. 319-346, and references cited therein).

15 In another embodiment, nucleic acid probes derived from the instant nucleic acid sequences may be used in direct fluorescence *in situ* hybridization (FISH) mapping (Trask (1991) *Trends Genet.* 7:149-154). Although current methods of FISH mapping favor use of large clones (several to several hundred KB; *see* Laan et al. (1995) *Genome Res.* 5:13-20), improvements in sensitivity may allow performance of FISH mapping using shorter probes.

20 A variety of nucleic acid amplification-based methods of genetic and physical mapping may be carried out using the instant nucleic acid sequences. Examples include allele-specific amplification (Kazazian (1989) *J. Lab. Clin. Med.* 11:95-96), polymorphism of PCR-amplified fragments (CAPS; Sheffield et al. (1993) *Genomics* 16:325-332), allele-specific ligation (Landegren et al. (1988) *Science* 241:1077-1080), nucleotide extension reactions (Sokolov (1990) *Nucleic Acid Res.* 18:3671), Radiation Hybrid Mapping (Walter et al. (1997) *Nat. Genet.* 7:22-28) and Happy Mapping (Dear and Cook (1989) *Nucleic Acid Res.* 17:6795-6807). For these methods, the sequence of a nucleic acid fragment is used to design and produce primer pairs for use in the amplification reaction or in primer extension reactions. The design of such primers is well known to those skilled in the art. In methods
25 employing PCR-based genetic mapping, it may be necessary to identify DNA sequence differences between the parents of the mapping cross in the region corresponding to the instant nucleic acid sequence. This, however, is generally not necessary for mapping methods.
30

Loss of function mutant phenotypes may be identified for the instant cDNA clones
35 either by targeted gene disruption protocols or by identifying specific mutants for these genes contained in a maize population carrying mutations in all possible genes (Ballinger and Benzer (1989) *Proc. Natl. Acad. Sci USA* 86:9402-9406; Koes et al. (1995) *Proc. Natl. Acad. Sci USA* 92:8149-8153; Bensen et al. (1995) *Plant Cell* 7:75-84). The latter approach

may be accomplished in two ways. First, short segments of the instant nucleic acid fragments may be used in polymerase chain reaction protocols in conjunction with a mutation tag sequence primer on DNAs prepared from a population of plants in which Mutator transposons or some other mutation-causing DNA element has been introduced (see Bensen, *supra*). The amplification of a specific DNA fragment with these primers indicates the insertion of the mutation tag element in or near the plant gene encoding the instant polypeptides. Alternatively, the instant nucleic acid fragment may be used as a hybridization probe against PCR amplification products generated from the mutation population using the mutation tag sequence primer in conjunction with an arbitrary genomic site primer, such as that for a restriction enzyme site-anchored synthetic adaptor. With either method, a plant containing a mutation in the endogenous gene encoding the instant polypeptides can be identified and obtained. This mutant plant can then be used to determine or confirm the natural function of the instant polypeptides disclosed herein.

EXAMPLES

The present invention is further defined in the following Examples, in which all parts and percentages are by weight and degrees are Celsius, unless otherwise stated. It should be understood that these Examples, while indicating preferred embodiments of the invention, are given by way of illustration only. From the above discussion and these Examples, one skilled in the art can ascertain the essential characteristics of this invention, and without departing from the spirit and scope thereof, can make various changes and modifications of the invention to adapt it to various usages and conditions.

EXAMPLE 1

Composition of cDNA Libraries; Isolation and Sequencing of cDNA Clones

cDNA libraries representing mRNAs from various corn, rice, soybean, and wheat tissues were prepared. The characteristics of the libraries are described below.

TABLE 2

cDNA Libraries from Corn, Rice, Soybean, and Wheat

Library	Tissue	Clone
cbn2	Corn Developing Kernel Two Days After Pollination	cbn2.pk0039.c2
p0060	Transgenic Corn Leaf Expressing Gene M1C07 (Leucine-Rich Repeat), Family 3-B7, Approximately One Month After Planting in Green House*	p0060.coran66r
p0110	Corn (Stages V3/V4**) Leaf Tissue Minus Midrib Harvested 4 Hours, 24 Hours and 7 Days After Infiltration With Salicylic Acid, Pooled*	p0110.cgsmk69r
rlr24	Rice Leaf 15 Days After Germination, 24 Hours After Infection of Strain <i>Magaporthe grisea</i> 4360-R-62 (AVR2-YAMO); Resistant	rlr24.pk0088.f7

Library	Tissue	Clone
rlr6	Rice Leaf 15 Days After Germination, 6 Hours After Infection of Strain <i>Magaporthe grisea</i> 4360-R-62 (AVR2-YAMO); Resistant	rlr6.pk0029.d7
rlr6	Rice Leaf 15 Days After Germination, 6 Hours After Infection of Strain <i>Magaporthe grisea</i> 4360-R-62 (AVR2-YAMO); Resistant	rlr6.pk0083.f9
rr1	Rice Root of Two Week Old Developing Seedling	rr1.pk084.j16
rsl1n	Rice 15-Day-Old Seedling*	rsl1n.pk013.h14
rsr9n	Rice Leaf 15 Days After Germination Harvested 2-72 Hours Following Infection With <i>Magnaporthe grisea</i> (4360-R-62 and 4360-R-67)*	rsr9n.pk004.p5
scb1c	Soybean Embryogenic Suspension Culture Subjected to 4 Bombardments and Collected 12 Hours Later	scb1c.pk003.d8
scb1c	Soybean Embryogenic Suspension Culture Subjected to 4 Bombardments and Collected 12 Hours Later	scb1c.pk004.k24
ses9c	Soybean Embryogenic Suspension	ses9c.pk001.j23
sgc2c	Soybean Cotyledon 12-20 Days After Germination (Mature Green)	sgs2c.pk002.p14
src3c	Soybean 8 Day Old Root Infected With Cyst Nematode	srr3c.pk003.d10
wdk1c	Wheat Developing Kernel, 3 Days After Anthesis	wdk1c.pk009.j19
wle1n	Wheat Leaf From 7 Day Old Etiolated Seedling*	wle1n.pk0042.f7
wlk8	Wheat Seedlings 8 Hours After Treatment With Herbicide***	wlk8.pk0006.a4
wlm96	Wheat Seedlings 96 Hours After Inoculation With <i>Erysiphe graminis f. sp tritici</i>	wlm96.pk032.n8
wre1n	Wheat Root From 7 Day Old Etiolated Seedling*	wre1n.pk164.b11

* These libraries were normalized essentially as described in U.S. Patent No. 5,482,845, incorporated herein by reference.

** Corn developmental stages are explained in the publication "How a corn plant develops" from the Iowa State University Coop. Ext. Service Special Report No. 48 reprinted June 1993.

*** Application of 6-iodo-2-propoxy-3-propyl-4(3*H*)-quinazolinone; synthesis and methods of using this compound are described in USSN 08/545,827, incorporated herein by reference.

cDNA libraries may be prepared by any one of many methods available. For example, the cDNAs may be introduced into plasmid vectors by first preparing the cDNA libraries in Uni-ZAP™ XR vectors according to the manufacturer's protocol (Stratagene Cloning Systems, La Jolla, CA). The Uni-ZAP™ XR libraries are converted into plasmid libraries according to the protocol provided by Stratagene. Upon conversion, cDNA inserts will be contained in the plasmid vector pBluescript. In addition, the cDNAs may be introduced directly into precut Bluescript II SK(+) vectors (Stratagene) using T4 DNA

ligase (New England Biolabs), followed by transfection into DH10B cells according to the manufacturer's protocol (GIBCO BRL Products). Once the cDNA inserts are in plasmid vectors, plasmid DNAs are prepared from randomly picked bacterial colonies containing recombinant pBluescript plasmids, or the insert cDNA sequences are amplified via
5 polymerase chain reaction using primers specific for vector sequences flanking the inserted cDNA sequences. Amplified insert DNAs or plasmid DNAs are sequenced in dye-primer sequencing reactions to generate partial cDNA sequences (expressed sequence tags or "ESTs"; see Adams et al., (1991) *Science* 252:1651-1656). The resulting ESTs are analyzed using a Perkin Elmer Model 377 fluorescent sequencer.

10 EXAMPLE 2

Identification of cDNA Clones

cDNA clones encoding proteinases were identified by conducting BLAST (Basic Local Alignment Search Tool; Altschul et al. (1993) *J. Mol. Biol.* 215:403-410; see also www.ncbi.nlm.nih.gov/BLAST/) searches for similarity to sequences contained in the
15 BLAST "nr" database (comprising all non-redundant GenBank CDS translations, sequences derived from the 3-dimensional structure Brookhaven Protein Data Bank, the last major release of the SWISS-PROT protein sequence database, EMBL, and DDBJ databases). The cDNA sequences obtained in Example 1 were analyzed for similarity to all publicly available DNA sequences contained in the "nr" database using the BLASTN algorithm
20 provided by the National Center for Biotechnology Information (NCBI). The DNA sequences were translated in all reading frames and compared for similarity to all publicly available protein sequences contained in the "nr" database using the BLASTX algorithm (Gish and States (1993) *Nat. Genet.* 3:266-272) provided by the NCBI. For convenience, the P-value (probability) of observing a match of a cDNA sequence to a sequence contained
25 in the searched databases merely by chance as calculated by BLAST are reported herein as "pLog" values, which represent the negative of the logarithm of the reported P-value. Accordingly, the greater the pLog value, the greater the likelihood that the cDNA sequence and the BLAST "hit" represent homologous proteins.

30 EXAMPLE 3

Characterization of cDNA Clones Encoding Calpain p94

The BLASTX search using the EST sequences from clones listed in Table 3 revealed similarity of the polypeptides encoded by the cDNAs to calpain p94 from *Homo sapiens* and *Drosophila melanogaster* (NCBI General Identifier Nos. 1345664 and 600420, respectively). Shown in Table 3 are the BLAST results for individual ESTs ("EST"):

TABLE 3

BLAST Results for Sequences Encoding Polypeptides
Homologous to Calpain p94

Clone	Status	NCBI General Identifier	BLAST pLog Score
cbn2.pk0039.c2	EST	1345664	12.52
rs11n.pk013.h14	EST	600420	11.40
ses9c.pk001.j23	EST	600420	31.70

The sequence of the entire cDNA insert in the clones mentioned above was determined. The BLASTP search using the amino acid sequences from clones listed in Table 4 revealed similarity of the polypeptides encoded by the cDNAs to calpain p94 from *Drosophila melanogaster* (NCBI General Identifier Nos. 600420 and 1079058). Shown in Table 4 are the BLAST results for the sequences of the entire cDNA inserts comprising the indicated cDNA clones ("FIS"):

TABLE 4

BLAST Results for Sequences Encoding Polypeptides
Homologous to Calpain p94

Clone	Status	BLAST pLog Score	
		600420	1079058
cbn2.pk0039.c2: fis	FIS	43.52	43.52
rs11n.pk013.h14: fis	FIS	43.00	43.00
ses9c.pk001.j23: fis	FIS	44.05	44.40

The data in Table 5 represents a calculation of the percent identity of the amino acid sequences set forth in SEQ ID NOs:2, 4, 6, 30, 32, and 34 and the *Drosophila melanogaster* sequences (NCBI General Identifier Nos. 600420 and 1079058).

TABLE 5

Percent Identity of Amino Acid Sequences Deduced From the Nucleotide
Sequences of cDNA Clones Encoding Polypeptides
Homologous to Calpain p94

SEQ ID NO.	Percent Identity to	
	600420	1079058
2	50.0	50.0
4	57.4	57.4
6	52.0	52.0
30	25.8	25.8
32	24.9	24.9
34	26.0	26.0

Sequence alignments and percent identity calculations were performed using the Megalign program of the LASERGENE bioinformatics computing suite (DNASTAR Inc., Madison, WI). Multiple alignment of the sequences was performed using the Clustal method of alignment (Higgins and Sharp (1989) *CABIOS*. 5:151-153) with the default parameters (GAP PENALTY=10, GAP LENGTH PENALTY=10). Default parameters for pairwise alignments using the Clustal method were KTUPLE 1, GAP PENALTY=3, WINDOW=5 and DIAGONALS SAVED=5. Sequence alignments and BLAST scores and probabilities indicate that the nucleic acid fragments comprising the instant cDNA clones encode a substantial portion of corn, rice and soybean calpain p94. These sequences represent the first plant sequences encoding calpain p94.

EXAMPLE 4

Characterization of cDNA Clones Encoding Cysteine Protease 1

The BLASTX search using the EST sequences from clones listed in Table 3 revealed similarity of the polypeptides encoded by the cDNAs to cysteine protease 1 from *Zea mays* (NCBI General Identifier No. 1706260). Shown in Table 6 are the BLAST results for individual ESTs ("EST"), or for contigs assembled from two or more ESTs ("Contig"):

TABLE 6
BLAST Results for Sequences Encoding Polypeptides
Homologous to Cysteine Protease 1

Clone	Status	BLAST pLog Score 1706260
rr1.pk084.j16	EST	94.52
Contig of: wdk1c.pk009.j19 wre1n.pk164.b11	Contig	130.00

The entire cDNA insert in clones rr1.pk084.j16 and wdk1c.pk009.j19 was determined. The BLASTP search using the amino acid sequences from clones listed in Table 7 revealed similarity of the polypeptides encoded by the cDNAs to cysteine protease 1 from *Zea mays* (NCBI General Identifier No. 1706260). Shown in Table 7 are the BLAST results for the sequences of the entire cDNA inserts comprising the indicated cDNA clones ("FIS"):

TABLE 7
BLAST Results for Sequences Encoding Polypeptides
Homologous to Cysteine Protease 1

Clone	Status	BLAST pLog Score 1706260
rr1.pk084.j16:fis	FIS	158.00
wdk1c.pk009.j19:fis	FIS	110.00

Amino acid sequence alignments using the Clustal method of alignment indicates that the rice sequence starts 88 amino acids down stream from the corn starting methionine, and that the wheat sequence starts 163 amino acids down stream from the corn starting methionine. The corn sequence has a signal sequence (amino acids 1-19) and a mature protein which corresponds to amino acids 137 through 371. Thus, the rice and wheat sequences included here contain the entire mature protein. The data in Table 8 represents a calculation of the percent identity of the amino acid sequences set forth in SEQ ID NOs:8, 10, 36, and 38 and the *Zea mays* sequence (NCBI General Identifier No. 1706260).

TABLE 8

Percent Identity of Amino Acid Sequences Deduced From the Nucleotide Sequences of cDNA Clones Encoding Polypeptides Homologous to Cysteine Protease 1

SEQ ID NO.	Percent Identity to 1706260
8	88.0
10	87.1
36	90.9
38	86.3

Sequence alignments and percent identity calculations were performed using the Megalign program of the LASERGENE bioinformatics computing suite (DNASTAR Inc., Madison, WI). Multiple alignment of the sequences was performed using the Clustal method of alignment (Higgins and Sharp (1989) *CABIOS*. 5:151-153) with the default parameters (GAP PENALTY=10, GAP LENGTH PENALTY=10). Default parameters for pairwise alignments using the Clustal method were KTUPLE 1, GAP PENALTY=3, WINDOW=5 and DIAGONALS SAVED=5. Sequence alignments and BLAST scores and probabilities indicate that the nucleic acid fragments comprising the instant cDNA clones encode a substantial portion of a rice and a wheat cysteine protease 1. These sequences represent the first rice and wheat sequences encoding cysteine protease 1.

EXAMPLE 5

Characterization of cDNA Clones Encoding Cysteine Protease 2

The BLASTX search using the EST sequences from clones listed in Table 9 revealed similarity of the polypeptides encoded by the cDNAs to cysteine protease 2 from *Phaseolus vulgaris* (NCBI General Identifier No. 2511691). Shown in Table 9 are the BLAST results for sequences of contigs assembled from two or more ESTs ("Contig"):

TABLE 9
BLAST Results for Sequences Encoding Polypeptides
Homologous to Cysteine Protease 2

Clone	Status	BLAST pLog Score 2511691
Contig of: sgs2c.pk002.p14 srr3c.pk003.d10 scb1c.pk003.d8	Contig	97.70

The sequence of the entire cDNA insert in clone srr3c.pk003.d10 was determined. The BLASTP search using the amino acid sequences from clones listed in Table 10 revealed similarity of the polypeptides encoded by the cDNAs to cysteine protease 2 from *Phaseolus vulgaris* (NCBI General Identifier No. 2511691). Shown in Table 10 are the BLAST results for the sequences of the entire cDNA inserts comprising the indicated cDNA clones encoding the entire protein ("CGS"):

TABLE 10
BLAST Results for Sequences Encoding Polypeptides
Homologous to Cysteine Protease 2

Clone	Status	BLAST pLog Score 2511691
srr3c.pk003.d10:fis	CGS	169.00

The data in Table 11 represents a calculation of the percent identity of the amino acid sequences set forth in SEQ ID NOs:12 and 40 and the *Phaseolus vulgaris* sequence (NCBI General Identifier No. 2511691).

TABLE 11
Percent Identity of Amino Acid Sequences Deduced From the Nucleotide
Sequences of cDNA Clones Encoding Polypeptides
Homologous to Cysteine Protease 2

SEQ ID NO.	Percent Identity to 2511691
12	67.9
40	75.1

Sequence alignments and percent identity calculations were performed using the Megalign program of the LASERGENE bioinformatics computing suite (DNASTAR Inc., Madison, WI). Multiple alignment of the sequences was performed using the Clustal method of alignment (Higgins and Sharp (1989) *CABIOS*. 5:151-153) with the default

parameters (GAP PENALTY=10, GAP LENGTH PENALTY=10). Default parameters for pairwise alignments using the Clustal method were KTUPLE 1, GAP PENALTY=3, WINDOW=5 and DIAGONALS SAVED=5. Sequence alignments and BLAST scores and probabilities indicate that the nucleic acid fragments comprising the instant cDNA clones encode a substantial portion and an entire soybean cysteine protease 2. These sequences represent the first soybean sequences encoding cysteine protease 2.

EXAMPLE 6

Characterization of cDNA Clones Encoding CLP Protease ATP-Binding Subunit

The BLASTX search using the EST sequences from clones listed in Table 12 revealed similarity of the polypeptides encoded by the cDNAs to two different CLP Protease ATP-binding subunits from *Lycopersicon esculentum* (NCBI General Identifier Nos. 399213 and 399212). Shown in Table 12 are the BLAST results for individual ESTs ("EST"), or for the sequences of contigs assembled from two or more ESTs ("Contig"):

TABLE 12
BLAST Results for Sequences Encoding Polypeptides
Homologous to CLP Protease ATP-Binding Subunit

Clone	Status	BLAST pLog Score	
		399213	399212
p0110.cgsmk69r	EST	66.52	66.00
Contig of: rlr6.pk0083.f9 rlr24.pk0088.f7 rlr6.pk0029.d7	Contig	126.00	124.00
wlm96.pk032.n8	EST	98.52	98.70

The sequence of the entire cDNA insert in clones p0110.cgsmk69r, rlr24.pk0088.f7, and wlm96.pk032.n8 was determined. The BLASTP search using the amino acid sequences from clones listed in Table 13 revealed similarity of the polypeptides encoded by the cDNAs to two different CLP Protease ATP-binding subunits from *Lycopersicon esculentum* (NCBI General Identifier Nos. 399213 and 399212). Shown in Table 13 are the BLAST results for the sequences of the entire cDNA inserts comprising the indicated cDNA clones ("FIS"):

TABLE 13

BLAST Results for Sequences Encoding Polypeptides
Homologous to CLP Protease ATP-Binding Subunit

Clone	Status	BLAST pLog Score	
		399213	399212
p0110.cgsmk69r: fis	FIS	>254.00	>254.00
rlr24.pk0088.f7: fis	FIS	>254.00	>254.00
wlm96.pk032.n8: fis	FIS	140.00	134.00

- 5 The data in Table 14 represents a calculation of the percent identity of the amino acid sequences set forth in SEQ ID NOs:14, 16, 18, 42, 44, and 46 and the *Lycopersicon esculentum* sequences (NCBI General Identifier Nos. 399213 and 399212).

TABLE 14

Percent Identity of Amino Acid Sequences Deduced From the Nucleotide
Sequences of cDNA Clones Encoding Polypeptides
Homologous to CLP Protease ATP-Binding Subunit

SEQ ID NO.	Percent Identity to	
	399213	399212
14	92.2	92.2
16	80.9	79.9
18	86.5	87.1
42	86.9	86.0
44	90.8	89.6
46	88.7	87.4

- 15 Sequence alignments and percent identity calculations were performed using the Megalign program of the LASERGENE bioinformatics computing suite (DNASTAR Inc., Madison, WI). Multiple alignment of the sequences was performed using the Clustal method of alignment (Higgins and Sharp (1989) *CABIOS*. 5:151-153) with the default parameters (GAP PENALTY=10, GAP LENGTH PENALTY=10). Default parameters for pairwise alignments using the Clustal method were KTUPLE 1, GAP PENALTY=3,
- 20 WINDOW=5 and DIAGONALS SAVED=5. Sequence alignments and BLAST scores and probabilities indicate that the nucleic acid fragments comprising the instant cDNA clones encode a substantial portion of corn, rice, and soybean CLP protease ATP binding domain. These sequences represent the first corn, rice, and soybean sequences encoding CLP protease ATP binding domain.

EXAMPLE 7

Characterization of cDNA Clones Encoding

CLP Protease Proteolytic Subunit

The BLASTX search using the EST sequences from clones listed in Table 15 revealed similarity of the polypeptides encoded by the cDNAs to CLPP from *Synechococcus PCC7942* (NCBI General Identifier No. 3023500), *Myxococcus xanthus* (NCBI General Identifier No. 3023519), or *Synechocystis sp.* (NCBI General Identifier No. 1705930). Shown in Table 15 are the BLAST results for individual ESTs ("EST"):

TABLE 15

BLAST Results for Sequences Encoding Polypeptides
Homologous to CLP Protease Proteolytic Subunit

Clone	Status	NCBI General Identifier No.	BLAST pLog Score
p0060.coran66r	EST	3023500	36.70
rsr9n.pk004.p5	EST	3023519	43.40
scb1c.pk004.k24	EST	1705930	60.52
wle1n.pk0042.f7	EST	2493737	45.70
wlk8.pk0006.a4	EST	3023519	43.22

The sequence of the entire cDNA insert in the clones mentioned above was determined. The BLASTP search using the amino acid sequences from clones listed in Table 16 revealed similarity of the polypeptides encoded by the cDNAs to two different CLPP from *Arabidopsis thaliana* (NCBI General Identifier Nos. 5360593 and 4887543) or from *Azospirillum brasilense* (NCBI General Identifier No. 6685315). Shown in Table 16 are the BLAST results for the sequences of the entire cDNA inserts comprising the indicated cDNA clones ("FIS"):

TABLE 16

BLAST Results for Sequences Encoding Polypeptides
Homologous to CLP Protease Proteolytic Subunit

Clone	Status	NCBI General Identifier No.	BLAST pLog Score
p0060.coran66r: fis	FIS	5360593	83.00
rsr9n.pk004.p5: fis	FIS	6685315	40.10
scb1c.pk004.k24: fis	FIS	4887543	103.00
wle1n.pk0042.f7: fis	FIS	4887543	86.00
wlk8.pk0006.a4: fis	FIS	6685315	54.05

The data in Table 17 represents a calculation of the percent identity of the amino acid sequences set forth in SEQ ID NOs:20, 22, 24, 26, 28, 48, 50, 52, 54, and 56 and the

Arabidopsis thaliana and *Azospirillum brasilense* sequences (NCBI General Identifier Nos. 5360593, 4887543, and 6685315).

TABLE 17

5 Percent Identity of Amino Acid Sequences Deduced From the Nucleotide Sequences of cDNA Clones Encoding Polypeptides Homologous to CLP Protease Proteolytic Subunit

SEQ ID NO.	Percent Identity to		
	5360593	4887543	6685315
20	78.8	52.2	53.1
22	44.0	55.0	73.4
24	48.7	97.3	58.7
26	55.8	91.6	63.2
28	57.1	64.3	81.0
48	78.8	39.7	38.6
50	34.0	43.8	57.6
52	45.5	92.5	51.5
54	39.0	47.1	60.5
56	48.9	87.9	54.6

10 Sequence alignments and percent identity calculations were performed using the Megalign program of the LASERGENE bioinformatics computing suite (DNASTAR Inc., Madison, WI). Multiple alignment of the sequences was performed using the Clustal method of alignment (Higgins and Sharp (1989) *CABIOS*. 5:151-153) with the default parameters (GAP PENALTY=10, GAP LENGTH PENALTY=10). Default parameters for pairwise alignments using the Clustal method were KTUPLE 1, GAP PENALTY=3, 15 WINDOW=5 and DIAGONALS SAVED=5. Sequence alignments and BLAST scores and probabilities indicate that the nucleic acid fragments comprising the instant cDNA clones encode a substantial portion of corn, rice, soybean, and wheat CLPPs. These sequences represent the first corn, rice, soybean, and wheat sequences encoding CLPP.

EXAMPLE 8

20 Expression of Chimeric Genes in Monocot Cells

25 A chimeric gene comprising a cDNA encoding the instant polypeptides in sense orientation with respect to the maize 27 kD zein promoter that is located 5' to the cDNA fragment, and the 10 kD zein 3' end that is located 3' to the cDNA fragment, can be constructed. The cDNA fragment of this gene may be generated by polymerase chain reaction (PCR) of the cDNA clone using appropriate oligonucleotide primers. Cloning sites (NcoI or SmaI) can be incorporated into the oligonucleotides to provide proper orientation of the DNA fragment when inserted into the digested vector pML103 as described below.

Amplification is then performed in a standard PCR. The amplified DNA is then digested with restriction enzymes NcoI and SmaI and fractionated on an agarose gel. The appropriate band can be isolated from the gel and combined with a 4.9 kb NcoI-SmaI fragment of the plasmid pML103. Plasmid pML103 has been deposited under the terms of the Budapest Treaty at ATCC (American Type Culture Collection, 10801 University Blvd., Manassas, VA 20110-2209), and bears accession number ATCC 97366. The DNA segment from pML103 contains a 1.05 kb SalI-NcoI promoter fragment of the maize 27 kD zein gene and a 0.96 kb SmaI-SalI fragment from the 3' end of the maize 10 kD zein gene in the vector pGem9Zf(+) (Promega). Vector and insert DNA can be ligated at 15°C overnight, essentially as described (Maniatis). The ligated DNA may then be used to transform *E. coli* XL1-Blue (Epicurian Coli XL-1 Blue™; Stratagene). Bacterial transformants can be screened by restriction enzyme digestion of plasmid DNA and limited nucleotide sequence analysis using the dideoxy chain termination method (Sequenase™ DNA Sequencing Kit; U.S. Biochemical). The resulting plasmid construct would comprise a chimeric gene encoding, in the 5' to 3' direction, the maize 27 kD zein promoter, a cDNA fragment encoding the instant polypeptides, and the 10 kD zein 3' region.

The chimeric gene described above can then be introduced into corn cells by the following procedure. Immature corn embryos can be dissected from developing caryopses derived from crosses of the inbred corn lines H99 and LH132. The embryos are isolated 10 to 11 days after pollination when they are 1.0 to 1.5 mm long. The embryos are then placed with the axis-side facing down and in contact with agarose-solidified N6 medium (Chu et al. (1975) *Sci. Sin. Peking* 18:659-668). The embryos are kept in the dark at 27°C. Friable embryogenic callus consisting of undifferentiated masses of cells with somatic proembryoids and embryoids borne on suspensor structures proliferates from the scutellum of these immature embryos. The embryogenic callus isolated from the primary explant can be cultured on N6 medium and sub-cultured on this medium every 2 to 3 weeks.

The plasmid, p35S/Ac (obtained from Dr. Peter Eckes, Hoechst Ag, Frankfurt, Germany) may be used in transformation experiments in order to provide for a selectable marker. This plasmid contains the *Pat* gene (see European Patent Publication 0 242 236) which encodes phosphinothricin acetyl transferase (PAT). The enzyme PAT confers resistance to herbicidal glutamine synthetase inhibitors such as phosphinothricin. The *pat* gene in p35S/Ac is under the control of the 35S promoter from Cauliflower Mosaic Virus (Odell et al. (1985) *Nature* 313:810-812) and the 3' region of the nopaline synthase gene from the T-DNA of the Ti plasmid of *Agrobacterium tumefaciens*.

The particle bombardment method (Klein et al. (1987) *Nature* 327:70-73) may be used to transfer genes to the callus culture cells. According to this method, gold particles (1 µm in diameter) are coated with DNA using the following technique. Ten µg of plasmid DNAs are added to 50 µL of a suspension of gold particles (60 mg per mL). Calcium chloride

(50 μ L of a 2.5 M solution) and spermidine free base (20 μ L of a 1.0 M solution) are added to the particles. The suspension is vortexed during the addition of these solutions. After 10 minutes, the tubes are briefly centrifuged (5 sec at 15,000 rpm) and the supernatant removed. The particles are resuspended in 200 μ L of absolute ethanol, centrifuged again and the supernatant removed. The ethanol rinse is performed again and the particles resuspended in a final volume of 30 μ L of ethanol. An aliquot (5 μ L) of the DNA-coated gold particles can be placed in the center of a Kapton™ flying disc (Bio-Rad Labs). The particles are then accelerated into the corn tissue with a Biolistic™ PDS-1000/He (Bio-Rad Instruments, Hercules CA), using a helium pressure of 1000 psi, a gap distance of 0.5 cm and a flying distance of 1.0 cm.

For bombardment, the embryogenic tissue is placed on filter paper over agarose-solidified N6 medium. The tissue is arranged as a thin lawn and covered a circular area of about 5 cm in diameter. The petri dish containing the tissue can be placed in the chamber of the PDS-1000/He approximately 8 cm from the stopping screen. The air in the chamber is then evacuated to a vacuum of 28 inches of Hg. The macrocarrier is accelerated with a helium shock wave using a rupture membrane that bursts when the He pressure in the shock tube reaches 1000 psi.

Seven days after bombardment the tissue can be transferred to N6 medium that contains glufosinate (2 mg per liter) and lacks casein or proline. The tissue continues to grow slowly on this medium. After an additional 2 weeks the tissue can be transferred to fresh N6 medium containing glufosinate. After 6 weeks, areas of about 1 cm in diameter of actively growing callus can be identified on some of the plates containing the glufosinate-supplemented medium. These calli may continue to grow when sub-cultured on the selective medium.

Plants can be regenerated from the transgenic callus by first transferring clusters of tissue to N6 medium supplemented with 0.2 mg per liter of 2,4-D. After two weeks the tissue can be transferred to regeneration medium (Fromm et al. (1990) *Bio/Technology* 8:833-839).

EXAMPLE 9

Expression of Chimeric Genes in Dicot Cells

A seed-specific expression cassette composed of the promoter and transcription terminator from the gene encoding the β subunit of the seed storage protein phaseolin from the bean *Phaseolus vulgaris* (Doyle et al. (1986) *J. Biol. Chem.* 261:9228-9238) can be used for expression of the instant polypeptides in transformed soybean. The phaseolin cassette includes about 500 nucleotides upstream (5') from the translation initiation codon and about 1650 nucleotides downstream (3') from the translation stop codon of phaseolin. Between the 5' and 3' regions are the unique restriction endonuclease sites Nco I (which includes the ATG

translation initiation codon), Sma I, Kpn I and Xba I. The entire cassette is flanked by Hind III sites.

The cDNA fragment of this gene may be generated by polymerase chain reaction (PCR) of the cDNA clone using appropriate oligonucleotide primers. Cloning sites can be incorporated into the oligonucleotides to provide proper orientation of the DNA fragment when inserted into the expression vector. Amplification is then performed as described above, and the isolated fragment is inserted into a pUC18 vector carrying the seed expression cassette.

Soybean embryos may then be transformed with the expression vector comprising sequences encoding the instant polypeptides. To induce somatic embryos, cotyledons, 3-5 mm in length dissected from surface sterilized, immature seeds of the soybean cultivar A2872, can be cultured in the light or dark at 26°C on an appropriate agar medium for 6-10 weeks. Somatic embryos which produce secondary embryos are then excised and placed into a suitable liquid medium. After repeated selection for clusters of somatic embryos which multiplied as early, globular staged embryos, the suspensions are maintained as described below.

Soybean embryogenic suspension cultures can be maintained in 35 mL liquid media on a rotary shaker, 150 rpm, at 26°C with fluorescent lights on a 16:8 hour day/night schedule. Cultures are subcultured every two weeks by inoculating approximately 35 mg of tissue into 35 mL of liquid medium.

Soybean embryogenic suspension cultures may then be transformed by the method of particle gun bombardment (Klein et al. (1987) *Nature* (London) 327:70-73, U.S. Patent No. 4,945,050). A DuPont Biolistic™ PDS1000/HE instrument (helium retrofit) can be used for these transformations.

A selectable marker gene which can be used to facilitate soybean transformation is a chimeric gene composed of the 35S promoter from Cauliflower Mosaic Virus (Odell et al. (1985) *Nature* 313:810-812), the hygromycin phosphotransferase gene from plasmid pJR225 (from *E. coli*; Gritz et al. (1983) *Gene* 25:179-188) and the 3' region of the nopaline synthase gene from the T-DNA of the Ti plasmid of *Agrobacterium tumefaciens*. The seed expression cassette comprising the phaseolin 5' region, the fragment encoding the instant polypeptides and the phaseolin 3' region can be isolated as a restriction fragment. This fragment can then be inserted into a unique restriction site of the vector carrying the marker gene.

To 50 µL of a 60 mg/mL 1 µm gold particle suspension is added (in order): 5 µL DNA (1 µg/µL), 20 µL spermidine (0.1 M), and 50 µL CaCl₂ (2.5 M). The particle preparation is then agitated for three minutes, spun in a microfuge for 10 seconds and the supernatant removed. The DNA-coated particles are then washed once in 400 µL 70% ethanol and resuspended in 40 µL of anhydrous ethanol. The DNA/particle suspension can

be sonicated three times for one second each. Five μ L of the DNA-coated gold particles are then loaded on each macro carrier disk.

Approximately 300-400 mg of a two-week-old suspension culture is placed in an empty 60x15 mm petri dish and the residual liquid removed from the tissue with a pipette.

For each transformation experiment, approximately 5-10 plates of tissue are normally bombarded. Membrane rupture pressure is set at 1100 psi and the chamber is evacuated to a vacuum of 28 inches mercury. The tissue is placed approximately 3.5 inches away from the retaining screen and bombarded three times. Following bombardment, the tissue can be divided in half and placed back into liquid and cultured as described above.

Five to seven days post bombardment, the liquid media may be exchanged with fresh media, and eleven to twelve days post bombardment with fresh media containing 50 mg/mL hygromycin. This selective media can be refreshed weekly. Seven to eight weeks post bombardment, green, transformed tissue may be observed growing from untransformed, necrotic embryogenic clusters. Isolated green tissue is removed and inoculated into individual flasks to generate new, clonally propagated, transformed embryogenic suspension cultures. Each new line may be treated as an independent transformation event. These suspensions can then be subcultured and maintained as clusters of immature embryos or regenerated into whole plants by maturation and germination of individual somatic embryos.

EXAMPLE 10

Expression of Chimeric Genes in Microbial Cells

The cDNAs encoding the instant polypeptides can be inserted into the T7 *E. coli* expression vector pBT430. This vector is a derivative of pET-3a (Rosenberg et al. (1987) *Gene* 56:125-135) which employs the bacteriophage T7 RNA polymerase/T7 promoter system. Plasmid pBT430 was constructed by first destroying the EcoR I and Hind III sites in pET-3a at their original positions. An oligonucleotide adaptor containing EcoR I and Hind III sites was inserted at the BamH I site of pET-3a. This created pET-3aM with additional unique cloning sites for insertion of genes into the expression vector. Then, the Nde I site at the position of translation initiation was converted to an Nco I site using oligonucleotide-directed mutagenesis. The DNA sequence of pET-3aM in this region, 5'-CATATGG, was converted to 5'-CCCATGG in pBT430.

Plasmid DNA containing a cDNA may be appropriately digested to release a nucleic acid fragment encoding the protein. This fragment may then be purified on a 1% NuSieve GTG™ low melting agarose gel (FMC). Buffer and agarose contain 10 μ g/ml ethidium bromide for visualization of the DNA fragment. The fragment can then be purified from the agarose gel by digestion with GELase™ (Epicentre Technologies) according to the manufacturer's instructions, ethanol precipitated, dried and resuspended in 20 μ L of water. Appropriate oligonucleotide adapters may be ligated to the fragment using T4 DNA ligase (New England Biolabs, Beverly, MA). The fragment containing the ligated adapters can be

purified from the excess adapters using low melting agarose as described above. The vector pBT430 is digested, dephosphorylated with alkaline phosphatase (NEB) and deproteinized with phenol/chloroform as described above. The prepared vector pBT430 and fragment can then be ligated at 16°C for 15 hours followed by transformation into DH5 electrocompetent cells (GIBCO BRL). Transformants can be selected on agar plates containing LB media and 100 µg/mL ampicillin. Transformants containing the gene encoding the instant polypeptides are then screened for the correct orientation with respect to the T7 promoter by restriction enzyme analysis.

For high level expression, a plasmid clone with the cDNA insert in the correct orientation relative to the T7 promoter can be transformed into *E. coli* strain BL21(DE3) (Studier et al. (1986) *J. Mol. Biol.* 189:113-130). Cultures are grown in LB medium containing ampicillin (100 mg/L) at 25°C. At an optical density at 600 nm of approximately 1, IPTG (isopropylthio-β-galactoside, the inducer) can be added to a final concentration of 0.4 mM and incubation can be continued for 3 h at 25°. Cells are then harvested by centrifugation and re-suspended in 50 µL of 50 mM Tris-HCl at pH 8.0 containing 0.1 mM DTT and 0.2 mM phenyl methylsulfonyl fluoride. A small amount of 1 mm glass beads can be added and the mixture sonicated 3 times for about 5 seconds each time with a microprobe sonicator. The mixture is centrifuged and the protein concentration of the supernatant determined. One µg of protein from the soluble fraction of the culture can be separated by SDS-polyacrylamide gel electrophoresis. Gels can be observed for protein bands migrating at the expected molecular weight.

Various modifications of the invention in addition to those shown and described herein will be apparent to those skilled in the art from the foregoing description. Such modifications are also intended to fall within the scope of the appended claims.

The disclosure of each reference set forth above is incorporated herein by reference in its entirety.